PENERAPAN METODE BAYES EMPIRIK PADA PENDUGAAN AREA KECIL UNTUK KASUS BINER (Studi tentang Proporsi Status Kepemilikan Kartu Sehat di Kota Yogyakarta)

Kismiantini
Jurusan Pendidikan Matematika
FMIPA Universitas Negeri Yogyakarta

ABSTRAK

Pendugaan area kecil berguna untuk menduga parameter subpopulasi (area) yang berukuran contoh kecil. Pada data biner, model Binomial-beta dapat digunakan untuk menduga parameter area kecil. Ada dua metode dalam pendugaan area kecil untuk data biner, yaitu Bayes empirik dan Bayes hierarkhi. Pada makalah ini digunakan metode Bayes empirik dalam menghasilkan penduga proporsi. Tujuannya adalah menilai kinerja penduga langsung dan penduga Bayes empirik pada pendugaan area kecil untuk kasus biner. Hasil penerapan pada proporsi status kepemilikan kartu sehat di kota Yogyakarta menunjukkan bahwa penduga Bayes empirik dari model Binomial-beta memberikan hasil pendugaan dengan ketelitian yang lebih tinggi dibandingkan penduga langsung.

Kata kunci: pendugaan area kecil, data biner, Bayes empirik

PENDAHULUAN

Pendugaan area kecil (*small area estimation*) adalah suatu teknik statistika untuk menduga parameter-parameter subpopulasi yang ukuran contohnya kecil (Rao, 2003a). Teknik pendugaan ini memanfaatkan data dari domain besar (yakni seperti data sensus, data survei sosial ekonomi nasional) untuk menduga peubah yang menjadi perhatian pada domain yang lebih kecil. Area kecil didefinisikan sebagai subpopulasi yang ukuran contohnya kecil sehingga pendugaan langsung tidak dapat menghasilkan dugaan yang teliti (Rao, 2003a). Biasanya statistik diperoleh dari suatu survei yang dirancang untuk memperoleh statistik nasional. Persoalan muncul ketika ingin diperoleh informasi untuk area yang lebih kecil (propinsi, kabupaten, kecamatan atau desa/kelurahan) yaitu objek survei jumlahnya kecil bahkan bisa saja area tersebut tidak tersampling sehingga analisis yang didasarkan hanya pada objek-objek tersebut menjadi sangat tidak dapat diandalkan (presisi rendah). *Small area estimation* merupakan suatu metode yang dapat menangani permasalahan tersebut.

Di Indonesia pentingnya statistik area kecil semakin dirasakan seiring dengan penerapan otonomi daerah, pemerintah daerah memiliki kewenangan yang lebih besar untuk mengatur dirinya sendiri. Kebutuhan statistik pada level daerah (kabupaten sampai kecamatan) menjadi keniscayaan sebagai dasar perencanaan pembangunan daerah atau kebijakan penting lainnya. Melalui metode ini, secara nasional akan cukup banyak biaya yang bisa dihemat oleh daerah sehingga dapat dialokasikan untuk pembiayaan pembangunan lainnya.

Berbagai metode pendugaan area kecil (*small area estimation*) telah dikembangkan khususnya menyangkut metode yang berbasis model (*model-based area estimation*). Metode tersebut adalah penduga prediksi tak bias linier terbaik empirik atau *empirical best linear unbiased prediction* selanjutnya disebut *EBLUP*, Bayes empirik atau *empirical Bayes* disingkat *EB*, dan Bayes hierarkhi atau *hierarchical Bayes* yang disingkat *HB*. Metode *EBLUP* merupakan metode untuk data kontinu sedangkan *EB* dan *HB* adalah metode untuk data biner atau cacahan. Metode Bayes empirik pada pendugaan area kecil pertama kali dilakukan oleh Fay dan Herriot (1979) dalam menduga pendapatan beberapa area kecil.

Suatu peubah respons yang menyatakan "sukses" atau "gagal" disebut sebagai peubah biner. Pada pendugaan area kecil untuk kasus biner, peubah yang menjadi perhatian berupa proporsi. Penduga langsung bagi proporsi merupakan penduga kemungkinan maksimum yaitu $\hat{p}_i = y_i / n_i$, dengan mengasumsikan peubah pengamatan diasumsikan menyebar binomial, $y_i \overset{iid}{\sim} Binomial(n_i, p_i)$. Penduga langsung ini mempunyai ragam yang besar karena hanya berdasarkan jumlah objek survei yang terdapat pada area tersebut. Suatu pendugaan lain dikembangkan untuk mengatasi permasalahan ini, yaitu pendugaan tak langsung. Pendugaan tak langsung bagi proporsi diperoleh dari model Binomial-beta. Model ini mempunyai dua tahap, yaitu pada tahap pertama diasumsikan bahwa peubah yang menjadi perhatian $y_i \overset{iid}{\sim} Binomial(n_i, p_i)$, $y_i = 0, \cdots, n_i$, $0 < p_i < 1$, $i = 1, \cdots, m$. Sedangkan pada tahap kedua diasumsikan bahwa $p_i \overset{iid}{\sim} beta(\alpha, \beta)$ sebagai prior, dengan fungsi kepadatan peluangnya adalah

$$f(p_i | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1} (1 - p_i)^{\beta-1}; \alpha > 0, \beta > 0 \qquad (1)$$

Berdasarkan teorema Bayes maka penduga Bayes dan ragam posterior bagi p_i adalah

$$\hat{p}_i^B(\alpha,\beta) = E(p_i | y_i, \alpha, \beta) = (y_i + \alpha)/(n_i + \alpha + \beta) \quad (2)$$

dan

$$Var(p_i | y_i, \alpha, \beta) = g_{1i}(\alpha, \beta, y_i) = \frac{(y_i + \alpha)(n_i - y_i + \beta)}{(n_i + \alpha + \beta + 1)(n_i + \alpha + \beta)^2} \quad (3)$$

Bila penduga Bayes ini akan digunakan maka harus diketahui terlebih dahulu nilai parameter sebaran priornya. Namun seringkali informasi mengenai parameter prior belum diketahui. Pendekatan lain yang dapat digunakan adalah Bayes empirik, yaitu pendekatan yang dilakukan untuk mendapatkan informasi parameter prior berdasarkan datanya. Informasi parameter prior diperoleh dengan memaksimumkan fungsi sebaran marjinal $y_i | \alpha, \beta \overset{ind}{\sim}$ Beta-binomial, namun bentuk tertutup untuk $\hat{\alpha}_{ML}$ dan $\hat{\beta}_{ML}$ tidak ada (McCulloch & Searle, 2001). Pada makalah ini bertujuan untuk menilai kinerja penduga langsung dan penduga Bayes empirik pada pendugaan area kecil untuk kasus biner.

PENDUGA LANGSUNG BAGI PROPORSI

Peubah pengamatan diasumsikan mempunyai sebaran binomial, $y_i \overset{iid}{\sim} Binomial(n_i, p_i)$. Fungsi peluang dari sebaran binomial adalah

$$f(y_i | p_i) = \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i}, \text{ dengan } y_i = 0, \cdots, n_i, \; 0 < p_i < 1, \; i = 1, \cdots, m \quad (4)$$

Selanjutnya dengan memaksimumkan fungsi peluang tersebut diperoleh penduga kemungkinan maksimum bagi p_i yaitu

$$\hat{p}_i = y_i / n_i \quad (5)$$

Penduga ini merupakan penduga kemungkinan maksimum yang bersifat tak bias karena nilai harapan dari penduga sama dengan parameternya.

$$E(\hat{p}_i) = E\left(\frac{y_i}{n_i}\right) = \frac{1}{n_i} E(y_i) = \frac{1}{n_i} n_i p_i = p_i \quad (6)$$

Sehingga dugaan kuadrat tengah galat sama dengan ragamnya, yaitu

$$ktg(\hat{p}_i) = \hat{V}ar(\hat{p}_i) = n_i \hat{p}_i (1 - \hat{p}_i) \quad (7)$$

PENDUGA BAYES EMPIRIK BAGI PROPORSI

Langkah awal pada pendugaan Bayes empirik dari model Beta-binomial oleh Kleinman (Rao, 2003a) adalah dengan membuat dugaan parameter prior dengan menyamakan rataan contoh terboboti

$$\hat{p} = \sum_i (n_i / n_T) \hat{p}_i \quad (8)$$

dan ragam contoh terboboti

$$s_p^2 = \sum_i (n_i / n_T)(\hat{p}_i - \hat{p})^2 \quad (9)$$

dengan nilai harapan masing-masing dan kemudian diselesaikan persamaan momen untuk α dan β, dengan $n_T = \sum_i n_i$. Penduga momen, $\hat{\alpha}$ dan $\hat{\beta}$, diberikan sebagai berikut:

$$\frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}} = \hat{p} \quad (10)$$

dan

$$\frac{1}{\hat{\alpha} + \hat{\beta} + 1} = \frac{n_T s_p^2 - \hat{p}(1-\hat{p})(m-1)}{\hat{p}(1-\hat{p})\left[n_T - \sum_i n_i^2 / n_T - (m-1)\right]} \quad (11)$$

Lalu substitusikan penduga momen $\hat{\alpha}$ dan $\hat{\beta}$ ke dalam (2) untuk memperoleh penduga Bayes empirik bagi p_i yaitu

$$\hat{p}_i^{EB} = \hat{p}_i^{B}(\hat{\alpha}, \hat{\beta}) = \hat{\gamma}_i \hat{p}_i + (1 - \hat{\gamma}_i)\hat{p} \quad (12)$$

dengan $\hat{\gamma}_i = n_i / (n_i + \hat{\alpha} + \hat{\beta})$, $\hat{p}_i = y_i / n_i$ sebagai penduga langsung dari p_i, y_i dan n_i masing-masing menyatakan banyaknya pengamatan dan banyaknya suatu kasus, $\hat{p}$ adalah penduga sintetik atau sebagai penduga tak langsung (Rao, 2003a).

PROSEDUR PENDUGAAN PROPORSI DENGAN METODE BAYES EMPIRIK

Prosedur yang digunakan dalam menduga proporsi ada dua cara yaitu berdasarkan penduga langsung dan penduga Bayes empirik dari model Binomial-Beta yang diuraikan sebagai berikut :

A. Penduga langsung
   1. Menentukan penduga proporsi

      $$\hat{p}_i = y_i / n_i$$

      dengan y_i menyatakan banyaknya pengamatan suatu kasus pada subpopulasi ke-i, n_i menyatakan banyaknya individu pada subpopulasi ke-i. Subpopulasi ini dapat berupa kecamatan.
   2. Menentukan dugaan kuadrat tengah galat (KTG) yaitu $\mathrm{ktg}(\hat{\theta}_i) = n_i \hat{p}_i (1 - \hat{p}_i)$.
   3. Menentukan galat baku.
   4. Proses hitungan dilakukan dengan Microsoft Office Excel.

B. Penduga Bayes empirik berdasarkan model Binomial-Beta
   1. Menentukan nilai dugaan parameter $\hat{\alpha}$ dan $\hat{\beta}$ dari sebaran prior

      $$p_i \overset{iid}{\sim} beta(\alpha, \beta)$$
   2. Menentukan penduga Bayes empirik $\hat{p}_i^{EB}$.
   3. Menentukan kuadrat tengah galat dengan menggunakan metode Jackknife yaitu :
      - Anggap bahwa $\hat{p}_i^{EB} = k_i(\hat{p}_i, \hat{\alpha}, \hat{\beta})$, $\hat{p}_{i,-l}^{EB} = k_i(\hat{p}_i, \hat{\alpha}_{-l}, \hat{\beta}_{-l})$, lalu

        $$\hat{M}_{2i} = \frac{m-1}{m} \sum_{l=1}^{m} \left(\hat{p}_{i,-l}^{EB} - \hat{p}_i^{EB} \right)^2$$
      - Dengan mencari $\hat{\alpha}_{-l}$ dan $\hat{\beta}_{-l}$ yang merupakan penduga momen yang diperoleh dari data ke-l yang dihapus, maka dihitung

        $$\hat{M}_{1i} = g_{1i}(\hat{\alpha}, \hat{\beta}, y_i) - \frac{m-1}{m} \sum_{l=1}^{m} \left[g_{1i}(\hat{\alpha}_{-l}, \hat{\beta}_{-l}, y_i) - g_{1i}(\hat{\alpha}, \hat{\beta}, y_i) \right]$$
      - Penduga Jackknife bagi kuadrat tengah galat penduga Bayes empirik diberikan oleh

        $$\mathbf{ktg}_J(\hat{\theta}_i^{EB}) = \hat{M}_{1i} + \hat{M}_{2i}$$

4. Menentukan galat baku.
5. Proses hitungan dilakukan dengan SAS/IML versi 9.1 dan Microsoft Office Excel.

Perbandingan kebaikan dari kedua penduga proporsi (penduga langsung dan Bayes empirik dari model Binomial-beta) dengan melihat nilai galat baku.

PENERAPAN PADA DATA STATUS KEPEMILIKAN KARTU SEHAT

Data status kepemilikan kartu sehat diperoleh dari Survei Sosial Ekonomi Nasional (SUSENAS) 2003 dengan materi informasi berbasis rumahtangga. Data ini diambil dari 14 kecamatan di kota Yogyakarta. Peubah yang menjadi perhatian adalah proporsi status kepemilikan kartu sehat, peubah pengamatan y_i adalah jumlah rumahtangga pemilik kartu sehat pada kecamatan ke-*i*, dan n_i adalah jumlah rumahtangga pada kecamatan ke-*i*.

Tabel 1. Pendugaan Proporsi Status Kepemilikan Kartu Sehat

No	Kecamatan	n_i	y_i	Langsung		Bayes Empirik	
				Penduga	Galat baku	Penduga	Galat baku
1	Mantrijeron	117	1	0,009	0,996	0,022	0,006
2	Kraton	110	4	0,036	1,963	0,036	0,005
3	Mergangsan	180	7	0,039	2,594	0,038	0,004
4	Umbulharjo	307	4	0,013	1,987	0,019	0,004
5	Kotagede	128	8	0,063	2,739	0,050	0,007
6	Gondokusuman	270	9	0,033	2,950	0,034	0,004
7	Danurejan	129	7	0,054	2,573	0,046	0,006
8	Pakualaman	62	0	0,000	0,000	0,023	0,007
9	Gondomanan	57	6	0,105	2,317	0,059	0,011
10	Ngampilan	41	0	0,000	0,000	0,026	0,007
11	Wirobrajan	120	2	0,017	1,402	0,026	0,005
12	Gedongtengen	97	4	0,041	1,958	0,038	0,005
13	Jetis	161	6	0,037	2,403	0,037	0,004
14	Tegalrejo	202	13	0,064	3,488	0,054	0,007

Dari Tabel 1 dapat diperoleh informasi bahwa secara rata-rata banyak rumahtangga belum memiliki kartu sehat. Penduga langsung memberikan galat baku yang besar sehingga penduga mempunyai presisi yang rendah, hal ini disebabkan kecilnya ukuran contoh pada area yang menjadi perhatian. Sedangkan penduga

Bayes empirik memberikan hasil pendugaan dengan presisi meningkat yang ditunjukkan oleh kecilnya galat baku.

SIMPULAN DAN SARAN

Berdasarkan hasil penerapan dapat diambil simpulan dan saran sebagai berikut:

Simpulan

Penduga Bayes empirik dari model Binomial-beta lebih dapat diandalkan daripada penduga langsung, yang diperlihatkan dengan semakin kecilnya galat baku.

Saran

Perbaikan pendugaan Bayes empirik pada model Binomial-beta dapat dilakukan dengan memasukkan peubah penyerta ke dalam model.

DAFTAR PUSTAKA

Brackstone, G.J. 2002. Strategies and approach for small area statistics. *Survey Methodology*, 28 (2), 117-123.

Carlin, B.P., & Louis, T.A. 2000. *Bayes and empirical Bayes methods for data analysis*. New York: Chapman & Hall.

Farrel, P.J., MacGibbon, B., & Tomberlin, T.J. 1997. Empirical Bayes small-area estimation using logistic regression models and summary statistics. *Journal of Business & Economic Statistics*, 15 (1),101-108.

Farrell, P.J., MacGibbon, B., & Tomberlin, T.J. 1997. Bootstrap adjusments for empirical Bayes interval estimates of small-area proportions. *The Canadian Journal of Statistics*, 25 (1), 75-89.

Fay, R.E., & Herriot, R.A. 1979. Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74 (366), 269-277.

Gill, J. 2002. *Bayesian methods*: *a social and behavioral sciences approach*. Boca Raton: Chapman & Hall.

Gosh, M., & Rao, J.N.K. 1994. Small area estimation: an appraisal. *Statistical Science*, 9 (1), 55-76.

Jiang, J., & Lahiri, P. 2001. Empirical best prediction for small area inference with binary data. *Annals of the Institute of Statistical Mathematics*, 53 (2), 217-243.

Jiang, J., Lahiri, P., & Wan, S.M. 2002. A unified jackknife theory for empirical best prediction with M-estimation. *The Annals of Statistics*, 30 (6),1782-1810.

Larsen, M.D. 2003. Estimation of small-area proportions using covariates and survey data. *Journal of Statistical Planning and Inference*, 112 (2003), 89-98.

Malec, D., Sedransk, J., Moriarity, C.L., & LeClere, F.B. 1997. Small area inference for binary variables in the national health interview survey. *Journal of the American Statistical Association*, 92 (439), 815-826.

McCulloch, C.E., & Searle, S.R. 2001. G*eneralized linear and mixed models*. New York: Wiley.

Rao, J.N.K. 1999. Some recent advances in model-based small area estimation. *Survey Methodology*, 25 (2), 175-186.

Rao, J.N.K. 2003a. *Small area estimation.* New York: John Wiley and Sons.

Rao, J.N.K. 2003b. Some new developments in small area estimation. *Proceedings of the survey methods selection*. Diambil pada tanggal 22 Februari 2006, dari http://www.ssc.ca/survey/documents/SSC2003_J_rao.pdf